DOCUMENT RESUME

ED 422 380 TM 028 942

AUTHOR Lai, Morris K.; Matsumoto, Annette N.; Young, Donald B.; Dougherty, Barbara J.

TITLE A Norm-Referenced, Performance-Based Mathematics Test Proves To Be Better at Revealing Effects of a Student-Driven Algebra Curriculum.

PUB DATE 1998-04-16

NOTE 9p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Diego, CA, April 13-17, 1998).

PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.

DESCRIPTORS Achievement Gains; Achievement Tests; *Algebra; *High School Students; High Schools; *Mathematics Tests; *Norm Referenced Tests; *Performance Based Assessment; *Standardized Tests; Teaching Methods; Test Use

IDENTIFIERS Hawaii

ABSTRACT

The Hawaii Learning Project (HALP) has produced an Algebra I curriculum that stresses student learning through problem solving, communication, connections, development over time, and challenging tasks. The HALP curriculum is used by more than 16,000 students in 13 states. Scores on standardized algebra tests for HALP graduates have been about the same as for students who have gone through a more traditional algebra program, but teachers of HALP students have strongly suggested that their students were doing better than students they had taught with more traditional approaches. Whether a standardized, norm-referenced commercially available test would be sensitive enough to show growth on the part of students using the HALP curriculum was studied. The most promising test available was the Harcourt-Brace GOALS: A Performance Based Measure of Achievement, which also had the advantage of having national norms and being equated scale-wise to the Metropolitan Achievement Test. GOALS scores were obtained from 190 Algebra I HALP students in Hawaii and Mississippi. Results show that this commercial, norm-referenced standardized performance-based test can reveal large gains beyond normative expectation, even though virtually no gains were shown with a more traditional standardized norm-referenced test. It is concluded that to assess the effects of an algebra program that reflects the new paradigm of curriculum recently espoused by the National Council of Teachers of Mathematics, commonly used algebra tests may not be valid. A test like GOALS may better reflect achievement in student-driven curricula. (Contains six tables and five references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 422 380

A Norm-Referenced, Performance-Based Mathematics Test Proves to be Better at Revealing Effects of a Student-Driven Algebra Curriculum

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Morris Lai
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
1

Morris K. Lai, Annette N. Matsumoto, & Donald B. Young
Curriculum Research & Development Group,
University of Hawai'i

Barbara J. Dougherty
Pacific Resources for Education and Learning (PREL)

TM028942

Paper presented at the annual meeting of the American Educational Research Association
Session 37.21
16 April 1998, San Diego, CA

A Norm-Referenced, Performance-Based Mathematics Test Proves to be Better at Revealing Effects of a Student-Driven Algebra Curriculum

The purpose of this study was to determine whether a standardized, norm-referenced commercially available test would be sensitive enough to show growth on the part of students using a student-driven curriculum featuring a problem-solving approach to algebra in which students are regularly required to explain their thinking in the classroom.

Perspective

The Hawaii Algebra Learning Project (HALP) has produced an Algebra I curriculum (Matsumoto, Dougherty, Wada, & Rachlin, 1994) that stresses student learning through problem solving, communication, connections, development over time, and challenging tasks. The HALP curriculum is used by more than 16,000 students in 13 states. Previous studies have shown that HALP graduates scored on standardized algebra tests about the same as students who have gone through a more traditional algebra program. Feedback from HALP teachers, however, strongly suggested that their students were doing better than students they had taught using more traditional approaches to algebra.

A project-developed test, while representing a better fit between the curriculum and the assessment, would always have a taint of possible project bias. Furthermore, without valid norms, it would be virtually impossible to use such a test to determine whether any gains were beyond expectation. The other main alternative of using control groups would have major difficulties such as finding truly comparable classes not using the HALP curriculum.

Method

Conceptually, the solution was simple—find a commercially available test that was standardized and norm referenced but was capable of showing whether students can better communicate their mathematical thinking through writing. Practically, the task to find such an instrument was daunting, especially in the field of algebra, which is laden with traditional types of problems that are likely to appear on tests. In searching in test reviews such as *Test Critiques* (Keyser & Sweetland, 1994) and *The Eleventh Mental Measurements Yearbook* (Kramer & Conoley, 1992), we were not able to locate algebra tests other than those using a multiple-choice format.

We then turned to mathematics tests not designed specifically for algebra. The most promising was Harcourt-Brace's *GOALS™: A Performance-Based Measure of*

Achievement, designed to be "a response to the demand that classroom assessment mirror more closely the kinds of instruction that students receive on a daily basis" (pg. 5). The test's open-ended format "assesses the integration of content and process necessary in today's curriculum" (pg. 5). Because *GOALS* emphasizes justification and explanation for answers, students must demonstrate their thinking and reasoning (Harcourt-Brace, 1994).

Not only did *GOALS* seem promising from a curriculum-fit viewpoint, it also had national norms and was equated scale-wise to the *Metropolitan Achievement Tests* (*MAT7*) and the *Stanford 8 (SAT8)*. Thus the test had the potential to be used to assess learning using sound designs such as (a) pre-post norm-referenced design or (b) posttest with *MAT7* or *SAT8* as a pretest or covariate. Although it is not a true control group, the national norming group provides an acceptable comparison group for statistical analyses.

Because the test scoring uses rubrics, it was necessary to have scorers (in this case, two mathematics teachers, who were not part of the project) undergo formal training (provided by the test publisher). High inter-rater reliability coefficients (greater than 90% exact agreement and greater than 95% exact or off by at most 1 on the 0–3 scale) verified that the training had been effective. Reliability checks conducted twice more during the scoring procedure verified that the high level of reliability had been maintained.

We arranged to have the *GOALS* test administered in fall and spring to HALP students in three widely differing sites. Two sites were in Mississippi, and one was in Hawai'i. While the Mississippi sites included both White and Black students, the Hawai'i site included an ethnic mix of students proportionally representative of the diverse population of the State.

In addition to the obvious ethnic differences between the Mississippi and Hawai'i sites, there were large differences in mathematics pre-levels of achievement. Mean pretest scores at the three sites corresponded to the 37th, 50th, and 71st (individual) percentiles. Complete data were collected from 190 students. All scoring of the tests was done blind as to whether the tests were pre or post.

To compare the scores, after computing means of the raw scores, we converted the means to their corresponding scaled scores. These scaled scores each corresponded to a percentile whose value depended on whether the test was administered in the fall or the spring.

Results

At all sites, large gains beyond normative expectation (see Table 1) were found (normative gains would have resulted in no changes in percentiles). Corresponding pre-post

percentiles were as follows (all were statistically significant at $p < .001$): Mississippi Site 1, 37th percentile pre to 54th percentile post; Mississippi Site 2, 50th percentile pre to 71st percentile post; Hawai'i Site, 71st percentile pre to 86th percentile post. A somewhat remarkable finding was that, even though there were large differences in pretest means at the three sites, the gains shown at each site were very similar in magnitude (between 15 and 21 percentile points), indicating a significant value-added component.

Table 1

Pre-Post Raw Scores and Corresponding Percentiles by Site

Site	Pretest Mean	Posttest Mean	Pretest Percentile	Posttest Percentile	Statistical Significance Level
1 (MS) $n = 95$	9.1 ($SD = 3.7$)	13.2 ($SD = 6.4$)	37	54	***
2 (MS) $n = 46$	11.5 ($SD = 4.7$)	16.6 ($SD = 7.0$)	50.5	71	***
3 (HI) $n = 49$	16.0 ($SD = 5.5$)	20.9 ($SD = 4.6$)	71	86.5	***

$N = 190$ students

$***p < .001$

Our subsequent investigation of race and gender subgroups turned up several interesting results. At Mississippi Site 1 (see Table 2), Black and White males scored on the pretest at exactly the same level, corresponding to the 35th percentile. On the posttest, White males were more than 10 percentile points higher than Black males, who themselves showed a gain of more than 10 percentile points from pre to post.

White females had pretest scores more than 14 percentile points higher than did Black females, with an even larger difference (27 percentile points) seen on the posttest. It should be noted that Black females also gained in percentile points beyond normative expectation. In the subgroup pre-post analyses in which race and gender were kept constant, all

differences at this site were statistically significant at $p < .05$ except for Black females ($p < .08$).

Table 2

Means and Corresponding Percentiles for Mississippi Site 1 Pretests and Posttests

Site 1		Pre Raw	Post Raw	Pre Percentile	Post Percentile	Percentile Gain
Site total	$N = 95$	9.1 ($SD = 3.7$)	13.2 ($SD = 6.4$)	37.5	54	16.5***
Black Males	$n = 15$	8.6 ($SD = 3.7$)	11.7 ($SD = 6.2$)	35	46.5	11.5*
White Males	$n = 28$	8.6 ($SD = 3.8$)	13.8 ($SD = 5.6$)	35	57	22***
Black Females	$n = 20$	8.0 ($SD = 3.3$)	10.2 ($SD = 6.3$)	31	38	7
White Females	$n = 32$	10.5 ($SD = 3.8$)	15.4 ($SD = 6.4$)	45.5	65	19.5***

*$p < .05$. ***$p < .001$.

At Mississippi Site 2 (see Table 3), White males scored noticeably higher on the pretest than did Black males, and White females scored noticeably higher on the posttest than did Black females. On the posttest, even though Black students gained on the average more than 12 percentile points beyond normative expectation, the White students gained even more and were therefore even further ahead of the Black students. In the subgroup pre-post analyses in which race and gender were kept constant, all differences at this site were statistically significant at $p < .05$ except for Black females ($p < .052$).

Table 3

Means and Corresponding Percentiles for Mississippi Site 2 Pretests and Posttests

Site 2		Pre Raw	Post Raw	Pre Percentile	Post Percentile	Percentile Gain
Site total	$N = 46$	11.5 ($SD = 4.7$)	16.6 ($SD = 7.0$)	50.5	71	20.5***
Black Males	$n = 9$	9.4 ($SD = 5.2$)	12.7 ($SD = 6.6$)	39	51.5	12.5*
White Males	$n = 10$	11.6 ($SD = 4.0$)	19.1 (SD = 7.3)	51	80.5	29.5**
Black Females	$n = 12$	9.8 ($SD = 4.8$)	13.2 ($SD = 5.2$)	41	54	13
White Females	$n = 15$	13.9 ($SD = 3.9$)	19.9 ($SD = 6.3$)	62.5	83.5	21**

*$p < .05$. **$p < .01$. ***$p < .001$.

The students from the Hawai'i site (see Table 4) were so ethnically diverse that ethnic comparisons would not have much meaning. At that site, females were slightly ahead of males on the pretest and about equal on the posttest, where their mean corresponded to a remarkable 86.5 percentile. In the subgroup pre-post analyses in which gender was kept constant, the differences at this site were statistically significant at $p < .001$.

Table 4

Means and Corresponding Percentiles for Hawai'i Pretests and Posttests

Site 3		Pre Raw	Post Raw	Pre Percentile	Post Percentile	Percentile Gain
Site total	$N = 49$	16.0 ($SD = 5.5$)	20.9 ($SD = 4.6$)	71.5	86.5	15***
Males	$n = 26$	15.4 ($SD = 5.9$)	21.0 ($SD = 4.7$)	69	87	18***
Females	$n = 23$	16.7 ($SD = 4.9$)	20.9 ($SD = 4.5$)	74	86.5	12.5***

***$p < .001$.

Conclusions

We have shown that a commercial standardized, norm-referenced performance-based test can reveal large gains beyond normative expectation, even though virtually no gains were shown on a more traditional standardized, norm-referenced test. The conclusion is clear: In order to properly assess the effects of an algebra program that reflects the new paradigms of curriculum such as recently espoused by the National Council of Teachers of Mathematics (1989), the commonly used algebra tests available commercially may not be valid. Whether the results would replicate in mathematics areas other than algebra needs to be investigated.

Those interested in evaluating curricula claiming to be constructivist or student driven should seriously consider investigating using tests like *GOALS*, which also has tests addressing reading, language, science, and social studies. If it turns out that such tests are successful in showing learning beyond expectation in cases where multiple-choice tests fail to show such a level of learning, then the field can use GOALS-like tests to become notably more knowledgeable about which programs as well as which instructional methods are effective.

References

Harcourt-Brace. (1994). *GOALS™: A performance-based measure of achievement—1992 multilevel norms book*. San Antonio, TX: Author.

Keyser, D. J., & Sweetland, R. C. (1994). *Test critiques*. Austin, TX: Pro-Ed.

Kramer, J. J., & Conoley, J. C. (Eds.). (1992). *The eleventh mental measurements yearbook*. Lincoln: University of Nebraska.

Matsumoto, A. N., Dougherty, B. J., Wada, L. A., & Rachlin, S. L. (1994). *Algebra I: A process approach—Teacher's guide*. Honolulu: University of Hawai'i, Curriculum Research & Development Group.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.